Examining the Validity and Reliability of a University's Teacher Performance Assessment (TPA)

Beverly M. Klecker

Morehead State University

Paper presented November 7, 2018 at the annual meeting of the Mid-South Educational Research Association. Pensacola, FL

Abstract

The Council for the Accreditation of Educator Preparation Programs (CAEP), required evidence of reliability and validity of measures used in a university's Educator Preparation Program (EPP). This paper describes processes that provided this evidence for the Teacher Performance Assessment (TPA). Literature examined included Messick (1989), Linn (1980). The TPA, a state-wide requirement, was introduced early in the university's EPP. Components were taught throughout coursework and clinical practice. Rubrics were aligned with Interstate Teacher Assessment and Support Consortium (InTASC) Model Core Teaching Standards and Learning Progressions for Teachers (CCSSO, 2017). Candidates were introduced to the TPA and rubrics in the Teacher Education Handbook. Course assignments were aligned with the TPA. Assessed knowledge and skills were taught for mastery. Accommodations were offered for candidates with documented learning differences. Candidates were given the opportunity throughout their work on the TPA to receive feedback and to revise sections. Three cycles of data were: Fall 2016: N=56; Spring 2017: N=94; and Fall 2017: N=56. TPAs for Cycles 1 and 2 were identical. The TPA for Cycle 3 had been revised. The EPP consistently interpreted data from the four-point evaluation scale as interval-level. Construct and content validity and reliability evidence: Factor analysis results: Cycle I: seven subscales explained 76.76% of the variance. Cycle 2 seven subscales explained 73.9% of the variance. Cronbach's alpha reliabilities for Cycle I and II subscales ranged from a high of Contextual Factors: α=.907; α=.921 to a low of Analysis of Student Learning α=.882; α=.897. Reliabilities for Cycle II were High: Reflection and Self-Evaluation α= .881.and low Contextual Factors α=.673. Inter-rater reliabilities, consistently high, were included and discussed in the study.

The Validity and Reliability of the Teacher Performance Assessment (TPA)

The Teacher Performance Assessment was selected for analyses of validity and reliability of measures because it is a comprehensive assessment that is introduced early in the program, is assessed through coursework, and is completed and evaluated during clinical practice late in the program. The rubrics for the TPA are aligned with InTASC Model Core Teaching Standards and Learning Progressions for Teachers (CCSSO, 2017). The performances are aligned with candidate requirements during clinical practice and are also aligned with the Kentucky Teacher Intern Performance (KTIP), the required in-field evaluation of first-year teachers in Kentucky.

Validity

Validity was defined by CAEP Glossary (2015) as:

> Validity**.** The extent to which a set of operations, test, or other assessment measures what it is supposed to measure. Validity is not a property of a data set but refers to the appropriateness of inferences from test scores or other forms of assessment and the credibility of the interpretations that are made concerning the findings of a measurement effort.

The following aspects of validity described by Messick (1995) were examined:

- **Content-related validity**: Do assessment items/components adequately and representatively sample the content area(s) to be measured?
- **Construct validity**: Do assessments and the assessment system measure the content they purport to measure?
- **Fairness**: Are all candidates afforded a fair opportunity to demonstrate their skills, knowledge, and dispositions?
- **Utility**: How useful are the data generated from assessments?
- **Prediction** (Criterion-related validity): How well do assessment instrument predict how well candidates will do in future situations?

- **Consequences**: Are assessment uses and interpretations contributing to increased candidate achievement and not producing unintended negative consequences? (Linn, 1994)

Content and Construct Validity Evidence

The content of the Techer Performance Assessment (TPA) is presented in the Teacher Education Program (TEP) Handbook online. All clinical practice students are required to complete a Teacher Performance Assessment (TPA). Compiling a TPA will provide you with professional growth experiences and documentation that reflects your ability to impact student learning. Working through the TPA design will assist you in developing "a teacher's way of thinking." This experience will prepare you for the Kentucky Teacher Internship Program (KTIP), which must be completed during your first year of teaching. (MSUTEPH, 2018).

The TPA was designed to measure seven skills. The means for the measures across three cycles are presented below. The number of items per scale varied across items. The scoring rubric contained four score points. The mean values of the scales were all between the scale values were 1 (lowest value) to 4 (highest value). The mean values on the scales were between the 3 and 4 score point (Table 1).

Table 1. Scale Means by Content Areas in Three Cycles of TPA Data

Teacher Performance Assessed	Fall 2016 N=56	Spring 2017 N=94	Fall 2017* N=56
Contextual Factors	3.24	3.42	3.67
Learning Objectives	3.41	3.68	3.42
Assessment Plan	3.28	3.44	3.21
Design for Instruction	3.32	3.55	3.47
Instructional Decision Making	3.32	3.48	Omitted
Analysis of Student Learning	3.24	3.48	3.41
Reflection and Self-Evaluation	3.14	3.47	3.28
Formatting Expectations	3.74	3.63	One Item:

*Fall 2017 Revised TPA

Use of Factor Analyses to Examine Structure of the TPA by Cycle

Factor analysis is a statistical technique that identifies the smaller number of factors/constructs/dimensions that underlie a larger set of variables (most of which are correlated to each other) (Sax, 1997). Factor analysis was used to examine the TPA data for each Cycle using SPSS software. Seven subscales were identified in Cycles 1

and 2 data; six subscales were identified in Cycle 3 data. Cronbach's alpha internal-consistency reliabilities were computed for each subscale. Factor analysis was used to identify the number of components in the TPA.

The structure of the TPA belongs to the data that are collected each cycle and not to the TPA instrument. Reliability is never an attribute of a test. Thompson and Vacha-Haase (2000) clearly explain this principle in their article "Psychometrics is Data Metrics: The test is not reliable." (Thompson & Vacha-Haase, 2000). Each time an instrument is used to collect data for measurement, reliability coefficients must be computed with the data collected.

Table 2. Factor Analysis for Fall 2016 TPA Cycle I Data (N=56)

Initial Eigenvalues				**Extraction Sums of Squared Loadings**		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	18.662	53.319	53.319	18.662	53.319	53.319
2	2.098	5.995	59.314	2.098	5.995	59.314
3	1.783	5.095	64.408	1.783	5.095	64.408
4	1.588	4.536	68.944	1.588	4.536	68.944
5	1.084	3.096	72.040	1.084	3.096	72.040
6	.851	2.430	74.470	.851	2.430	74.470
7	.801	2.289	76.759	.801	2.289	76.759
8	.785	2.243	79.002			
9	.657	1.877	80.879			

Seven components were extracted explaining 76.759% of the variance in the Fall 2016 data. The analysis confirmed the seven designed scales. This presents statistical evidence of construct validity and content validity of the Fall 2016 Teacher Performance Assessment data. The TPA measured what it was designed to measure. The reader is reminded that "structure" and reliability do not belong to an instrument. Instrument structure and reliabilities belong to the data that were collected.

Cronbach alpha reliabilities are measures of internal consistency (Sax, 1997). Reliability coefficients do not "belong to a test," they belong to the data that were collected by an instrument. Thompson and Vacha-Hasse (2000) explained in "The Test is Not Reliable."

Cronbach alpha reliabilities are frequently used as evidence of both consistence in measurement (the usual definition for reliability) but also for construct and content

validity. With high internal consistency reliability, users are assured of both credible measurement and strong internal reliability. This is excellent confirmation of the underlying construct measuring the content.

Table 3. Factor Analysis for the Spring 2017 TPA Cycle 2 Data (N= 94)

Initial Eigenvalues				Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.889	49.060	49.060	6.889	49.060	49.060
2	.855	6.086	55.146	.855	6.086	55.146
3	.735	5.234	60.380	.735	5.234	60.380
4	.587	4.184	64.564	.587	4.184	64.564
5	.515	3.668	68.232	.515	3.668	68.232
6	.425	3.026	71.258	.425	3.026	71.258
7	.370	2.638	73.895	.370	2.638	73.895
8	.316	2.252	76.148			
9	.282	2.011	78.159			
10	.272	1.936	80.095			

Through factor analysis, seven components were identified in the Spring 2017 TPA Data. The seven components explained 73.9% of the variance in the data (Table 3 above). This presents statistical evidence of construct validity and content validity using the Spring 2017 Teacher Performance Assessment data. The TPA measured what it was designed to measure.

Table 4. Factor Analysis for the Fall 2017 TPA Data: N=56 (SPSS)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.701	37.018	37.018	5.701	37.018	37.018
2	1.552	10.076	47.095	1.552	10.076	47.095
3	1.299	8.433	55.528	1.299	8.433	55.528
4	.843	5.477	61.005	.843	5.477	61.005
5	.747	4.854	65.859	.747	4.854	65.859
6	.617	4.007	69.866	.617	4.007	69.866
7	.541	3.511	73.377			
8	.434	2.820	76.198			
9	.425	2.759	78.956			
10	.347	2.252	81.208			

The Teacher Performance Assessment was revised prior to the Fall 2017 semester. The purpose given for the revision was to clarify and simplify the rubrics. The Instructional Decision-Making scale was omitted in the Fall 2017 TPA.

Only one item was used to replace "Format Expectations" in the first two cycles. This one item was an evaluation of References. This one item did not constitute a "scale". In the analysis of the data, six scales were identified through factor analysis with SPSS. These six scales accounted for 69.86% of the variance, a decrease of variance explained from Fall 2016 (76.75%) and Spring 2017 (73.89%).

Table 5. TPA Cronbach Alpha Subscale Reliabilities across Cycles

Teacher Performance Assessed	Fall 2016 N=56 α	Spring 2017 N=94 α	Fall 2017 N=56 α
Contextual Factors	5 items α=.907	5 items α=.921	5 items α=.673**
Learning Goals/Objectives	4 items α= .887	4 items α=.850	4 items α=870
Assessment Plan	5 items α=.903	5 items α=.892	5 items α=.821
Design for Instruction	6 items α=.924	6 items α=882	4 items α=.838
Instructional Decision Making	3 items α=.875	3 items α=834	Omitted
Analysis of Student Learning	4 items α=.882	4 items α=897	4 items α=.757**
Reflection and Self-Evaluation	5 items α=.930	5 items α=.875	6 items α=.881
Formatting Expectations*	3 items α=.683**	3 items α=.690**	One item only "References"

*Not counted as a subscale. **Reliability below acceptable α=.800

CAEP's value expectation for acceptable reliabilities is at least .800 (CAEP, 2015). Cronbach's alpha measures reliability by the internal consistency of the items, that is, each of the items is measuring the same construct or content. Cronbach's alpha is used in the development and evaluation of reliability of instrument scales and subscales.

Seven components were identified in the data through factor analysis. These subscales in each year of data are: Contextual Factors, Learning Goals/Objectives, Assessment Plan, Design for Instruction, Instructional Decision Making (omitted in Fall 2017), Analysis of Student Learning, and Reflection and Self-Evaluation. "Formatting Expectations" was not defined in this study as a subscale.

The TPA data from the first two Cycles (Fall 2016, and Spring 2017) had seven identified subscales: Contextual Factors, Learning Goals/Objectives, Assessment Plan,

Design for Instruction, Instructional Decision Making, Analysis of Student Learning, and Reflection and Self-evaluation. The subscales were confirmed through factor analysis and each has a Cronbach's alpha internal consistency of .80 or higher. "Formatting Expectations" was not defined as a subscale in this study.

The TPA data from the third Cycle (Fall 2017) resulted from a revised TPA for Fall 2017. Six subscales were confirmed in the Fall TPA data through factor analysis: Contextual Factors, Learning Goals/Objectives, Assessment Plan, Design for Instruction, Analysis of Student Learning, and Reflection and Self-evaluation.

Reliability coefficients for two of the subscales in the Fall 2017 TPA data were below CAEP acceptable reliability levels: Contextual Factors (α=.673) and Analysis of Student Learning (α=.757). Results from the two subscales should NOT be used for making decisions about a candidate's knowledge and skills of in these two areas. Nor should the results from these two subscales be used in the aggregate. The lack of reliability in the subscales should be of concern to decision-makers. A review of data across Cycles should be helpful in making corrections to these "measures".

Cronbach's alpha, as other measures of reliability, is a statistical calculation. The more variance in the data, the higher the measure of internal consistency (true of any correlational computation). For assessments, a "rule of thumb" is to have at least 25 items with five points in the rating scale or four multiple-choice items for classroom assessments or research surveys. Thus, the more items used to measure a TPA scale, the higher the internal-consistency reliability. Alternatively, returning to the measures used in Cycles I and II should be helpful.

Validity: Fairness and the Opportunity to Learn (OTL) the Requirements of the TPA

Candidates are introduced to the Teacher Performance Assessment in the Teacher Education Handbook. Each year's handbook includes a copy of the TPA and Rubrics. Required courses have assignments aligned with the required sections of the TPA **(see Appendix for course alignments with the TPA for three Cycles).** The knowledge and skills are taught for mastery in the classes. Accommodations are offered in each class syllabus for candidates with documented learning differences.

Candidates have opportunities to demonstrate their learning within classrooms using feedback from university and clinical faculty. Candidates are given the opportunity throughout their work on the TPA to receive feedback and to re-do sections.

Reliability

CAEP Definition of Reliability and Inter-Rater Reliability (CAEP Glossary, 2015):

The degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable and repeatable for an individual test taker. A measure is said to have a high reliability if it produces consistent results under consistent conditions. Tables 6 and 7 below present Inter-Rater Agreement (Graham, Milanowski, & Miller, February 2012) for Cycles 1, 2, and 3. IRA provides exact scale agreement by at least two raters.

Table 6. Inter-Rater Agreement for Teacher Performance Evaluations: Cycles 1 and 2

Standards	CYCLE 1: Fall 2016 N=56				CYCLE 2: Spring 2017 N=94			
Contextual Factors	**IRA %***	**% Rated 3 or 4****	**MEAN**	**MODE**	**IRA %***	**% Rated 3 or 4****	**MEAN**	**MODE**
Knowledge of Community, School and Classroom Factors	81.1%	98.1%	3.37	3.00	73.7%	95.0%	3.60	4.00
Knowledge of Characteristics of Students	69.8%	94.3%	3.23	3.00	63.1%	93.0%	3.51	4.00
Knowledge of Students' Varied Approaches to Learning	86.7%	88.7%	3.24	3.00	57.4%	81.9%	3.27	4.00
Knowledge of Students' Skills and Prior Learning	84.9%	86.8%	3.27	3.00	62.7%	81.9%	3.40	4.00
Implications for Instructional Planning and Assessment	90.6%	84.90%	3.16	3.00	70.2%	86.7%	3.33	4.00
Learning Goals	**IRA %**	**% Rated 3 or 4****	**MEAN**	**MODE**	**IRA % ***	**% Rated 3 or 4****	**MEAN**	**MODE**
Significance, Challenge and Variety	83.0%	88.7%	3.31	4.00	74.4%	95.2%	3.68	4.00
Clarity	90.6%	98.1%	3.47	4.00	77.6%	95.2%	3.70	4.00
Appropriateness for Students	88.7%	98.1%	3.38	3.00	76.6%	94.8%	3.65	4.00
Alignment with National, State or Local Standards	96.2%	96.2%	3.50	4.00	77.7%	94.8%	3.70	4.00

*IRA=Percent of TPAs with Rater Exact Score Agreement (Scale range 1-4). **Percent of TPAs rated 3 or 4 (Adjacent Score).

Table 6. Inter-Rater Agreement for Teacher Performance Evaluations: Cycles 1 and 2 (Continued)

Standards	**CYCLE 1: Fall 2016 N=56**				**CYCLE 2: Spring 2017 N=94**			
Assessment Plan	**IRA %***	**% Rated 3 or 4****	**MEAN**	**MODE**	**IRA %***	**% Rated 3 or 4****	**MEAN**	**MODE**
Alignment with Learning Goals and Instruction	88.7%	92.55	3.33	3.00	64.9%	93.3%	3.49	4.00
Clarity of Criteria and Standards for Performance	86.8%	92.4%	3.34	4.00	71.3%	93.3%	3.54	4.00
Multiple Modes and Approaches	88.7%	92.4%	3.36	4.00	55.3%	92.4%	3.40	4.00
Technical Soundness	88.7%	92.4%	3.27	3.00	**NOT ASSESSED**			
Adaptations Based on the Individual Needs of Students	88.7%	81.1%	3.24	3.00	70.2%	84.9%	3.35	3.00
Design for Instruction	**IRA %**	**% Rated 3 or 4****	**MEAN**	**MODE**	**IRA % ***	**% Rated 3 or 4****	**MEAN**	**MODE**
Alignment with Learning Goals	88.7%	94.33%	3.37	4.00	63.5%	96.7%	3.67	4.00
Accurate Representation of Content	88.7%	94.33%	3.31	3.00	63.5%	95.7%	3.48	4.00
Lesson and Unit Structure	96.2%	92.4%	3.35	3.00	73.4%	95.7%	3.52	4.00
Use of a Variety of Instruction, Activities, Assignments and Resources	88.7%	86.8%	3.29	3.00	68.0%	88.7%	3.50	4.00
Use of Contextual Information and Data to Select Appropriate and Relevant Activities..	88.7%	88.7%	3.27	3.00	68.0%	88.7%	3.54	4.00
Use of Technology	86.8%	92.4%	3.47	4.00	72.3%	90.5%	3.56	4.00

Table 6. Inter-Rater Agreement for Teacher Performance Evaluations: Cycles 1 and 2 (Continued)

Standards	CYCLE 1: Fall 2016 N=53				CYCLE 2: Spring 2017 N=94			
Instructional Decision-Making	**IRA %***	**Rated 3 or 4**	**MEAN**	**MODE**	**IRA %***	**Rated 3 or 4**	**MEAN**	**MODE**
Sound Professional Practice	92.5%	92.4%	3.41	4.00	79.7%	94.6%	3.66	4.00
Modifications Based on Analysis of Student Learning	100%	92.5%	3.33	3.00	63.8%	92.4%	3.45	4.00
Congruence Between Modifications and Learning Goals	92.5%	88.7%	3.26	3.00	61.8%	91.4%	3.32	3.00
Analysis of Student Learning	**IRA %**	**Rated 3 or 4**	**MEAN**	**MODE**	**IRA % ***	**Rated 3 or 4**	**MEAN**	**MODE**
Clarity and accuracy of Presentation	75.5%	90.6%	3.34	3.00	67.0%	89.2%	3.56	4.00
Alignment with Learning Goals	84.9%	86.8%	3.31	3.00	67.0%	88.2%	3.54	4.00
Interpretation of Data	86.8%	84.9%	3.17	3.00	54.2%	84.9%	3.38	4.00
Evidence of Impact on Student Learning	86.8%	88.7%	3.20	3.00	62.7%	91.4%	3.38	4.00
Reflection and Self-Evaluation	**IRA %**	**Rated 3 or 4**	**MEAN**	**MODE**	**IRA % ***	**Rated 3 or 4**	**MEAN**	**MODE**
Interpretation of Student Learning	88.7%	86.8%	3.20	3.00	61.7%	89.2%	3.37	4.00
Insights on Effective Instruction and Assessment	86.8%	88.7%	3.13	3.00	58.5%	84.9%	3.41	4.00
Alignment Among Goals, Instruction and Assessment	84.9%	83.0%	3.20	3.00	63.8%	89.2%	3.37	4.00

*IRA=Percent of TPAs with Rater Exact Score Agreement (Scale range 1-4). **Percent of TPAs rated 3 or 4 (Adjacent Score).

Table 6. Inter-Rater Agreement for Teacher Performance Evaluations: Cycles 1 and 2 (Continued)

CYCLE 1: Fall 2016 N=56				CYCLE 2: Spring 2017 N=94				
Reflection and Self-Evaluation	**IRA %**	**Rated 3 or 4**	**MEAN**	**MODE**	**IRA % ***	**Rated 3 or 4**	**MEAN**	**MODE**
Implications for Future Teaching	88.7%	90.5%	3.15	3.00	53.2%	89.2%	3.39	4.00
Implications for Professional Development	79.2%	90.5	3.10	3.00	65.9%	91.4%	3.29	3.00
Format Expectations	**IRA %**	**Rated 3 or 4**	**MEAN**	**MODE**	**IRA % ***	**Rated 3 or 4**	**MEAN**	**MODE**
Format and Organization	94.3%	98.1%	3.85	4.00	73.4%	90.3%	3.69	4.00
Writing	86.8%	92.4%	3.76	4.00	62.7%	93.8%	3.44	4.00
References	98.1%	98.1%	3,94	4.00	77.6%	96.75	3.77	4.00

*IRA=Percent of TPAs with Rater Exact Score Agreement (Scale range 1-4). **Percent of TPAs rated 3 or 4 (Adjacent Score).

Table 7. Teacher Performance Assessment (TPA) Evaluations: Cycle 3 Fall 2017*

Standards	CYCLE 3: Fall 2017 N=56			
Contextual Factors	**IRA %***	**% Rated 3 or 4**	**MEAN**	**MODE**
The Setting for Learning - Community	73.2%	94.7%	3.78	4.00
The Setting for learning - School, Classroom and Students	71.4%	91.0%	3.71	4.00
The Setting for Learning	73.2%	98.2%	3.78	4.00
The Setting for Learning Selection of Monitored Students	60.7%	85.7%	3.51	4.00
The Setting for Learning Impact of Contextual Factors	64.2%	85.7%	3.61	4.00
Learning Objectives	**IRA % ***	**% Rated 3 or 4**	**MEAN**	**MODE**
Planning Sequence Organization	66.0%	76.8%	3.25	4.00
Planning Sequence Objectives	71.4%	75.0%	3.48	4.00
Planning for Content Alignment (National/State Standards)	66.1%	71.4%	3.41	4.00
Planning Academic Language	58.9%	76.7%	3.49	4.00
Assessment Plan	**IRA % ***	**% Rated 3 or 4**	**MEAN**	**MODE**
Planning Pre-Assessments	58.9%	69.6%	3.41	4.00
Planning Multiple Assessment Modes for Formative Assessment..	42.8%	71.4%	3.33	4.00
Planning Classroom Assessment with Student Learning	35.7%	64.2%	3.23	4.00
Planning Classroom Assessment Measures for Learners	39.3%	64.3%	3.11	3.00
Planning Classroom Assessment Based on Target Students Needs	32.17%	42.8%	2.97	3.00
Design for Instruction	**IRA % ***	**% Rated 3 or 4**	**MEAN**	**MODE**
Alignment of Learning Objectives	60.7%	26.8%	3.61	4.00
Engaging the Learner	57.1%	83.9%	3.45	4.00
Measuring Learning Growth of the Learner	46.1%	87.5%	3.49	4.00
Use of a Variety of Instruction, Activities, Assignments and Resources	41.1%	78.5%	3.34	4.00

NOTE: The TPA was revised prior to the Fall 2017 Assessment.

*IRA=Percent of TPAs with Rater Exact Score Agreement (Scale range 1-4).

**Percent of TPAs rated 3 or 4 (Adjacent Score).

Table 7. Teacher Performance Assessment (TPA) Evaluations: Cycle 3 Fall 2017* (Continued)

Standards	Cycle 3: Fall 2017 N=56			
Design for Instruction (continued)	**IRA % ***	**% Rated 3 or 4**	**MEAN**	**MODE**
Modifications Based on Learners' Needs	41.1%	78.5%	3.33	3.00
Analysis of Student Learning	**IRA % ***	**% Rated 3 or 4**	**MEAN**	**MODE**
Clarity and accuracy of Whole Group Learning	62.5%	98.2%	3.64	4.00
Impact of Whole Class Learning Feedback	42.8%	82.1%	3.40	4.00
Interpretation of Data of Target Learners	41.4%	82.1%	3.35	3.00
Evidence of Impact of Target Student	50.0%	75.0%	3.28	3.00
Reflection and Self-Evaluation	**IRA % ***	**% Rated 3 or 4**	**MEAN**	**MODE**
Interpretation of Student Learning	46.4%	67.8%	3.20	3.00
Teacher Candidate Insight on Effective Instruction and Assessment	41.0%	75.0%	3.23	3.00
Self-evaluation and Reflection to Improve Planning and Practice	46.40%	76.6%	3.24	3.00
Implications for Future Teaching	44.6%	60.7%	3.09	3.00
Format and Organization	71.4%	94.6	3.64	4.00
Writing	42.8%	89.2%	3.39	3.00
References	82.1%	94.6%	3.86	4.00

NOTE: The TPA was revised prior to the Fall 2017 Assessment.

*IRA=Percent of TPAs with Rater Exact Score Agreement (Scale range 1-4).

**Percent of TPAs rated 3 or 4 (Adjacent Score).

Interpretation of Reliability Tables for TPA Cycles 1, 2, and 3

The Tables present data for three cycles of the Teacher Performance Assessment (TPA). Cycle 1 is Fall 2016; Cycle 2 is Spring 2017; Cycle 3 is the "revised" TPA data from Fall 2017.

The data were analyzed using Excel. Note the variance in the IRA (Inter-Rater Agreement) by TPA task (Graham, Milanowski, & Miller, February, 2012). The IRA is exact score agreement, a measure that is now recommended for evaluations in education---especially in the ratings of teacher performance. The IRA replaces IRR--inter-rater reliability. IRR was often confusing, and calculation was dependent on the number of points on the rating scale used. IRA is a more transparent measure.

The IRA percentage is interpreted by the percentage of TPAs that received exactly the same rating. This is presented by individual "tasks" of TPA in order to provide instructional feedback for professors and candidates.

Note the IRA for selecting the two highest scores on the scale--indicate the percentage of TPA that were scored either a 3 or 4. Please note that the reason that the IRA (exact agreement) las lower percentages is that the majority of the "disagreement" in rating is apparently between the scores of 3 or 4.Care should be taken in future TPA scorer training sessions to spend enough time on helping scorers make clear distinctions between the ratings of "3" and "4."

References

Council for the Accreditation of Educator Preparation (CAEP). Glossary (2015). Retrieved from http://caepnet.org/glossary

Council of Chief State School Officers (CCSSO) (November, 2017). InTASC *Model Core Teaching Standards and Learning Progressions for Teachers 1.0.* Washington, DC. Retrieved from https://www.ccsso.org/resource-library/intasc-model-core-teaching-standards-and-learning-progressions-teachers-10

Graham, M., Milanowski, A., & Miller, J. (February, 2012). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings.* ERIC Document No. ED532068. Full text available. Retrieved from: https://eric.ed.gov/?id=ED532068

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749. Retrieved from: http://dx.doi.org/10.1037/0003-066X.50.9.741

Morehead State University Teacher Education Program Handbook (MSUTEPH). (2018). Teacher performance assessment and rubrics. Retrieved from http://www.moreheadstate.edu/College- of-Education/Teacher-Education-Services/Clinical-Practice/Teacher-Performance-Assessment

Sax, G. (1997). *Principles of educational and psychological measurement and evaluation*. Wadsworth Publishing Company.

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is data metrics: The test is not reliable. *Educational and Psychological Measurement, 60,* 174-195.

APPENDIX

Validity: Fairness and the Opportunity to Learn (OTL) Knowledge and Skills for the TPA MSU EPSB Key Assessments / CAEP 3-Cycle Assessments--Initial Certification Programs--Cycle/Course identification--08/ 2017 draft

	EPSB area #4 Observation			EPSB area #5 Assessment Plan			EPSB area #6 Contextual Factors			EPSB area #7 Lesson Plan			EPSB area #8 Literacy Assessment			EPSB area #9 Professional Attributes and Dispositions (Phil. Paper, Ethics, etc.)		
Program	**Cyc1 Course**	**Cyc2 Course**	**Cyc3 Course**	**Cyc1 Course**	**Cyc2 Course**	**Cyc3 Course**	**Cyc1 Course**	**Cyc2 Course**	**Cyc3 Course**	**Cyc1 Course**	**Cyc2 Course**	**Cyc3 Course**	**Cyc1 Course**	**Cyc2 Course**	**Cyc3 Course**	**Cyc1 Course**	**Cyc2 Course**	**Cyc3 Course**
IECE	IECE 361	IECE 418	IECE 425	IECE 255	EDSP 230	IECE 425	EDEC 253	IECE 301	IECE 425	EDEE 305	IECE 345	IECE 425	EDEE 327	EDSP 320	IECE 425	EDF 207	TEPADM	IECE 425
P-5	EDEM 330	EDEE 321	TPA	EDSP 230	EDSP 367	TPA	EDF 211	EDSP 365	TPA	EDEE 305	SCI 490 EDEE 322	TPA	EDEE 327	EDEE 331	TPA	EDF 207	TEPADM	TPA
P-5/ LBD P-12	EDEM 330	EDSP 359	TPA	EDSP 230	EDSP 367	TPA	EDF 211	EDSP 365	TPA	EDEE 305	SCI 490 EDEE 322	TPA	EDEE 327	EDEE 331	TPA	EDF 207	TEPADM	TPA
P-5 / MSD P-12	EDEM 330	EDSP 375	TPA	EDSP 230	EDSP 367	TPA	EDF 211	EDSP 365	TPA	EDEE 305	SCI 490 EDEE 322	TPA	EDEE 327	EDEE 331	TPA	EDF 207	TEPADM	TPA
MIDDLE	EDEM 330	EDMG 332	TPA	EDSP 230	EDUC 482	TPA	EDF 211	METH	TPA	EDMG 306	METH	TPA	EDMG 347	EDMG 332	TPA	EDF 207	TEPADM	TPA
MIDDLE / LBD P-12	EDEM 330	EDSP 359	TPA	EDSP 230	EDUC 482	TPA	EDF 211	EDSP 365	TPA	EDMG 306	EDSP 357	TPA	EDMG 347	EDMG 332		EDF 207	TEPADM	TPA
MIDDLE / MSD P-12	EDEM 330	EDSP 375	TPA	EDSP 230	EDSP 375 EDUC 482	TPA	EDF 211	EDSP 365	TPA	EDMG 306	METH	TPA	EDMG 347	EDMG 332	TPA	EDF 207	TEPADM	TPA
SOCIAL STUDIES 8-12	EDSE 451? EDSE 499D?	EDSE 451? EDSE 499D?	EDSE 416	EDSP 230	EDF 311	EDSE 416	EDF 211	EDSE 451? EDSE 499D?	EDSE 416	EDSE 451? EDSE 499D?	EDSE 451? EDSE 499D?	EDSE 416	EDUC 476			EDF 207	TEPADM	TPA
PHYSICS 8-12	SCI 402	SCI 402	EDSE 416	EDSP 230	EDF 311	EDSE 416	EDF 211	SCI 402	EDSE 416	SCI 402	SCI 402	EDSE 416	EDUC 476			EDF 207	TEPADM	TPA

	SCI 403	SCI 403						SCI 403		SCI 403	SCI 403					
CHEMISTRY 8-12	SCI 402	SCI 402	EDSE 416	EDSP 230	EDF 311	EDSE 416	EDF 211	SCI 402 SCI 403	EDSE 416	SCI 402 SCI 403	SCI 402 SCI 403	EDSE 416	EDUC 476	EDF 207	TEP ADM	TPA
BIOLOGY 8-12	SCI 402	SCI 402	EDSE 416	EDSP 230	EDF 311	EDSE 416	EDF 211	?	EDSE 416	?	?	EDSE 416	EDUC 476	EDF 207	TEP ADM	TPA
MATH 8-12	SCI 402	SCI 402	EDSE 416	EDSP 230	EDF 311	EDSE 416	EDF 211	?	EDSE 416	?	?	EDSE 416	EDUC 476	EDF 207	TEP ADM	TPA
AGRICULTURE 8-12	CTE 388 CTE 470	CTE 388 CTE 470	EDSE 416	EDSP 230	EDF 311	EDSE 416	EDF 211	CTE 388 CTE 470	EDSE 416	CTE 388 CTE 470	CTE 388 CTE 470	EDSE 416	EDUC 476	EDF 207	TEP ADM	TPA
ENGINEERING TECH 9-12	CTE 388 CTE 470	CTE 388 CTE 470	EDSE 416	EDSP 230	EDF 311	EDSE 416	EDF 211	CTE 388 CTE 470	EDSE 416	CTE 388 CTE 470	CTE 388 CTE 470	EDSE 416	EDUC 476	CTE 207	TEP ADM	TPA
SPANISH P-12	?	?	EDSE 416	EDSP 230	EDF 311	EDSE 416	EDF 211	?	EDSE 416	?	?	EDSE 416	EDUC 476	EDF 207	TEP ADM	TPA
MUSIC P-12	MUSE 215	MUSE 207	EDSE 416	EDSP 230	MUSE 375/ 376	EDSE 416	EDF 211	MUSE 375/ 376	EDSE 416	MUSE 207	MUSE 375/ 376 MUSE 325	EDSE 416	EDUC 476	EDF 207	TEP ADM	TPA
ART P-12	ART 300	ART 301 ART 321	EDSE 416	EDSP 230	?	EDSE 416	EDF 211	ART 300	EDSE 416	ART 300	ART 301 ART 321	EDSE 416	EDUC 476	EDF 207	TEP ADM	TPA
THEATRE P-12	THEA 375 THEA 370	THEA 475?	EDSE 416	EDSP 230	EDF 311	EDSE 416	EDF 211	THEA 375 THEA 370 THEA 475	EDSE 416	THEA 375 THEA 370	THEA 475	EDSE 416	EDUC 476	EDF 207	TEP ADM	TPA
ENGLISH 8-12	ENG 382	ENG 400	EDSE 416	EDSP 230	EDF 311	EDSE 416	EDF 211	ENG 382	EDSE 416	ENG 382	ENG 400	EDSE 416	EDUC 476 OR ENG 381, ENG 382, and EDSE 416	EDF 207	TEP ADM	TPA
BUSINESS INFO TECH 8-12	BIS 499C	BIS 499C	EDSE 416	EDSP 230	EDF 311	EDSE 416	EDF 211	BIS 499C	EDSE 416	BIS 499C	BIS 499C	EDSE 416	EDUC 476	EDF 207	TEP ADM	TPA